IMAGE POINTS
and
RIEMANN'S THEOREM

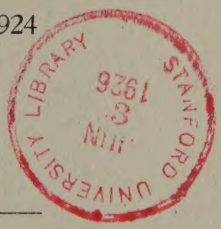
A DISSERTATION

SUBMITTED TO THE BOARD OF UNIVERSITY STUDIES
OF THE JOHNS HOPKINS UNIVERSITY
IN CONFORMITY WITH THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

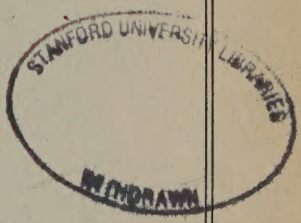
BY

FRANCIS JOSEPH GERST

1924



LOYOLA UNIVERSITY PRESS
CHICAGO, ILLINOIS
1925



3781
J65G
UNCLASSED
Dissertation

IMAGE POINTS
and
RIEMANN'S THEOREM

A DISSERTATION

SUBMITTED TO THE BOARD OF UNIVERSITY STUDIES
OF THE JOHNS HOPKINS UNIVERSITY
IN CONFORMITY WITH THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

BY

FRANCIS JOSEPH GERST

1924

LOYOLA UNIVERSITY PRESS
CHICAGO, ILLINOIS
1925

INTRODUCTION

Map-making is an ancient art. It existed, no doubt, among very primitive peoples, and had as its object the representation, on a small scale, of a portion of the earth's surface. When man began to investigate more closely the principles underlying the art, the science of map-making budded forth and developed. At first the navigator and geographer claimed this science as peculiarly their own; their right to it for practical purposes cannot be gainsaid.

With the varied development of mathematics in the last few centuries, many new phases of the ancient problem presented themselves. Through Lambert and Gauss to Riemann there is a steady advance toward a purely mathematical investigation of the problem of the conformal representation of one area upon another. The high-water mark in the theory is undoubtedly Riemann's fundamental existence theorem, that "there exists a function which will map conformally a region of one plane bounded by a curve onto a region in another plane bounded by a given curve."

Riemann's proof of this theorem was based on Dirichlet's principle: "It is evident that of all functions V, which with their first derivatives are continuous in a given closed region, and which take on assigned values at the boundary, there exists at least one for which the integral

$$\int\int\int\left[\left(\frac{\delta V}{\delta x}\right)^2+\left(\frac{\delta V}{\delta y}\right)^2+\left(\frac{\delta V}{\delta z}\right)^2\right]dx\,dy\,dz$$

taken over the whole region has a minimum value." Weierstrass uncovered some weak points in Riemann's proof, and soon a flood of literature on Dirichlet's principle inundated the mathematical world. The much-disputed principle and the fundamental theorem are now, however, firmly established.

In the meantime attempts were made to set Riemann's theorem on the firm foundation of pure function theory. The work of Schwarz, who was the first to distinguish between the mapping of inner points and boundary points, opened the way for the new method. Poincaré and others helped to advance it, while Koebe, Carathéodory, and Bieberbach were principally instrumental in bringing this method of proof to its present stage of completeness. Notably Koebe, by a series of important lemmas which he enunciated and proved, clearly defined the limits of the problem and reduced it to its simplest terms.

The present paper treats of certain phases of this existence theorem. The first part is an attempt to apply to it the methods of inversive geometry, in which the dominant idea is the self-inverse character of certain curves. The second part treats of the function-theory method of proof.

I. IMAGE POINTS

If we write an equation of the form $f(x, \bar{y}) = 0$, where x and y are complex numbers (and a bar over a quantity indicates the conjugate complex of that quantity), then every point x of the complex plane will have one or more correspondents, and, vice versa, every point y will determine one or more points x. If $x = y$, the equation becomes $f(x, \bar{x}) = 0$. Equations of this character are divided into two classes: (1) those which are self-conjugate, $f = \bar{f}$; (2) those which are not self-conjugate, $f \neq \bar{f}$.

When $x = y$, we call x a double point. If $f = \bar{f}$, we say that $f(x, \bar{x}) = 0$ is a curve, the locus of double points, and define as images in (or as to) this curve the x- and y-points of $f(x, \bar{y}) = 0$.

This last statement implies that an equation $f(x, \bar{x}) = 0$ is the analytic representation of a curve when and only when $f = \bar{f}$. The use of the word "curve" in this connection does not contradict the accepted notion of curve in the case of real functions of real variables. For let x and $\bar{x}$ be functions of rectangular Cartesian co-ordinates X and Y. If f is equal to its conjugate, then $f(x, \bar{x})$ is a real function of real variables, say $F(X, Y)$, and $F = 0$ is a curve. If f is not equal to its conjugate, then F is not real, since some of the coefficients are imaginary.

As an illustration of the foregoing, let

$$f(x, \bar{y}) \equiv \frac{x}{a} + \frac{\bar{y}}{\bar{a}} - 1 = 0 \tag{1}$$

where a and $\bar{a}$ are conjugate complex constants. Every x determines one y and vice versa. The points of the plane are paired off. If in (1) we replace $\bar{y}$ by $\bar{x}$, we get

$$f(x, \bar{x}) \equiv \frac{x}{a} + \frac{\bar{x}}{\bar{a}} - 1 = 0 \,. \tag{2}$$

In this case $f=\bar{f}$, and (2) is a locus. The points x and y are images of one another in (2), and (2) itself is the locus of double points or of coincident images. Equation (1) is the equation of image points. If $f(x, \bar{x})$ is not self-conjugate, it remains true that $f(x, \bar{y})=0$ pairs off the points of the plane, but there is no real locus of double points in which the paired-off points are "imaged."

Geometrically, if o is the base point and a its image in the line,

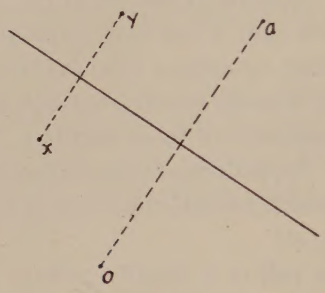


Fig. 1

and x and y images of each other in the same line, then the triangles xao and yoa are reversely similar. Hence

$$\begin{vmatrix} x & a & o \\ \bar{y} & o & \bar{a} \\ 1 & 1 & 1 \end{vmatrix} = 0 \quad \text{or} \quad \frac{x}{a}+\frac{\bar{y}}{\bar{a}}=1$$

If x is on the line, so is y, and $x=y$, and $\bar{x}=\bar{y}$. Hence

$$\frac{x}{a}+\frac{\bar{x}}{\bar{a}}=1$$

which is the equation of the line.

If $f(x, \bar{y})=0$ contains squares or higher powers, then every point will have more than one image. Consider the equation

$$(x-a)^2+(\bar{y}-\bar{a})^2=c\bar{c} \tag{3}$$

$$\bar{y}-\bar{a}=\pm\sqrt{c\bar{c}-(x-a)^2}\,. \tag{3'}$$

Each x determines two y's, y_1 and y_2, $\dfrac{y_1-a}{y_2-a}=-1$. Whence y_1 and y_2 are symmetric as to a. Now (3) is a rectangular hyperbola, center a; y_1 and y_2 are the images of x in this hyperbola. Hence the images in a rectangular hyperbola of any point are symmetric as to its center. If x runs to the hyperbola, y_1 will also, and coincide with x, and y_2 will run to the other branch. This result can also be deduced by using a Riemann surface. Let

$$w=z^2 \qquad w=u+iv \qquad z=X+iY\ .$$

The w-surface is two sheeted and has branch points at zero and infinity. Make the negative u-axis the branch cut. The two lines (one in each sheet) parallel to the v-axis map into the two branches of a rectangular hyperbola in the z-plane. The point x (Fig. 2) has

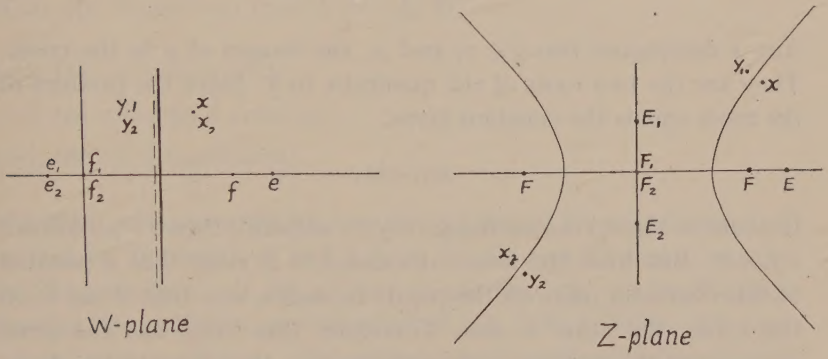


Fig. 2

two images, one in each line of the two sheets. These map into image points of x, in the hyperbola. Starting with a point x and finding the images, and then the images of these image points, not more than four points arise. Moreover, if one of the four is on the curve, all are on the curve. The point e has images e_1 and e_2. As e goes to f, e_1 and e_2 go to zero and there coincide, since zero is a branch point. In the z-plane, E goes to F, and E_1 and E_2 coincide at the origin (F_1 and F_2). The point F is a point two of whose images coincide, and is called a focus.

THE BIQUADRATIC IN x AND $\bar{x}$

If we assume a relation $xy=k^2$, (k any constant), the points of the plane will be paired off. Let $m+\bar{m}=1$ be a line which picks out the mid-points of certain pairs of points (x, y). Then

$$2m=x+y, \qquad \text{whence } 2\bar{m}=\bar{x}+\bar{y}$$

$$x+y+\bar{x}+\bar{y}=2 \text{ or } \qquad x+\frac{k^2}{x}+\bar{x}+\frac{\bar{k}^2}{\bar{x}}=2$$

$$\text{i.e., } x\bar{x}(x+\bar{x}-2)+k^2\bar{x}+\bar{k}^2x=0$$

a biquadratic in x and $\bar{x}$. In projective nomenclature this is a cubic. It is a self-conjugate equation, hence it represents a curve. To get the images of a point in the cubic, write $\bar{y}$ for $\bar{x}$. The equation of image points is

$$\bar{y}^2+\frac{x^2-2x+k^2}{x}\bar{y}+\bar{k}^2=0 .$$

Any x determines two y's, y_1 and y_2, the images of x in the cubic. They are the two roots of the quadratic in $\bar{y}$. Since the product of the roots equals the constant term,

$$y_1y_2=k^2 .$$

If x runs to the curve, one image, say y_1, will also; then $x=y_1$, whence $x\,y_2=k^2$. But from the above discussion it is clear that a relation of this character pairs off the points in such a way that if one is on the cubic, the other is also. Therefore, this cubic has the same property as the rectangular hyperbola, viz., that if a point and one of its images run to the curve, the second image is also on the curve. Curves which have this property are called self-inverse curves. The cubics found by varying k are the analogues among cubics in general of the rectangular hyperbola among conics. All conics are not self-inverse, nor are all cubics.

THE FOCI

A focus is a point two of whose images coincide. To find the foci, make $y_1=y_2$. This implies the vanishing of the discriminant. Hence

$$x^2-2x+k^2=\pm 2\bar{k}x .$$

The four foci, f_1, f_2, f_3, f_4, are the four roots of these two equations.

Suppose k real; then $k=\bar{k}$. For

$k<\frac{1}{2}$, the foci are real and distinct;

$k>\frac{1}{2}$, two foci real, two conjugate complexes;

$k=\frac{1}{2}$, two foci real and coincident, two real and distinct.

These self-inverse cubics can be mapped onto the line $z+\bar{z}=1$. The mapping equation is $2z=x+k^2/x$. The z-surface is two sheeted and has branch points at $\pm k$. The map of the line will be, in the x-plane, the cubic

$$x+\frac{k^2}{x}+\bar{x}+\frac{\bar{k}^2}{\bar{x}}=2.$$

Make the branch cut from k to $-k$. If

$k<\frac{1}{2}$, the cubic is bipartite;

$k=\frac{1}{2}$, the cubic has a crunode;

$k>\frac{1}{2}$, the cubic is unipartite.

Any point x has two images, one in each line of the two sheets, and conversely, the points y_1 and y_2 have images x at the same point in the distinct sheets. If the x's run to the line, the y's do likewise, showing the self-inverse character of the cubic. There are four points, two in each sheet, whose images are coincident (at the branch points k and $-k$). Hence, as above, four foci. If $k<\frac{1}{2}$ (the bipartite case), the foci are the images of the four points $1+k$ and $1-k$ (two in each sheet). If $k=\frac{1}{2}$, this point is also a (double) focus; the other foci are $f_3=\frac{3}{2}+1\ 2, f_4=\frac{3}{2}-1\ 2$ (maps of $z=\frac{3}{2}$, the image of $-k$). In the unipartite case, the algebra shows that two foci are real and two conjugate complexes. This appears likewise from the mapping. The four images of k and $-k$ lie on the real axis. One pair lies on the branch cut. From the mapping equation, $x=z+1\sqrt{z^2-k^2}$, when z is real, and equal to or greater than k, x is real. When z is real and less than k, the x-points are conjugate complexes. Table I gives the numerical calculation of the foci for four values of k.

TABLE I

k	f_2	f_3	f_4	f_5
1/4.....	0.04	2.47	0.03	1.46
31/64.....	0.34	2.8	0.07	0.69
1/2.....	0.5	2.9	0.09	0.5
1.....	$-i$	3.7	0.3	i

A well-known cubic which is self-inverse is the strophoid. Writing the general equation of the rectangular hyperbola

$$(x-a)^2+(\bar{x}-\bar{a})^2-c\bar{c}=0$$

and subjecting it to the transformation $x\bar{x}=a\bar{a}$, we get

$$\left(\frac{a\bar{a}}{\bar{x}}-a\right)^2+\left(\frac{a\bar{a}}{x}-\bar{a}\right)^2-c\bar{c}=0$$

which reduces to

$$(a^2+\bar{a}^2-c\bar{c})x^2\bar{x}^2+a^2\bar{a}^2(x^2+\bar{x}^2)-2a\bar{a}x\bar{x}(ax+\bar{a}\bar{x})=0\ .$$

If $a^2+\bar{a}^2=c\bar{c}$, the curve is the strophoid. Since coincidence of points and lines is an invariantive property under a homographic transformation, it follows that the strophoid is a self-inverse curve. Similarly, if the equation of the rectangular hyperbola is $x^2+\bar{x}^2=c\bar{c}$ and the transformation $x\bar{x}=1$, the resulting equation is

$$x^2+\bar{x}^2=c\bar{c}\ x^2\ \bar{x}^2$$

a lemniscate, which is likewise a self-inverse curve. Using the functional relation

$$w=\frac{1}{\bar{z}^2}\qquad w=u+iv\qquad \bar{z}=X-iY$$

the Riemann surface for w is two sheeted and has branch points at zero and infinity. The line $u=k$ of the two-sheeted w-surface maps onto the lemniscate

$$k(X^2+Y^2)^2=X^2-Y^2\ .$$

As in the case of the rectangular hyperbola, so analogously the proof of the self-inverse character of the lemniscate proceeds step by step. Although it is true that when the map equation is $w=1/\bar{z}^2$, w is not a function of the (standard) complex variable, yet it is a function of a complex variable, the "conjugate," and the mapping has all the characteristics of the usual mapping except that angles are reversed.[1]

F_1 $(=1/\sqrt{2k})$ has two coincident images at $z=\infty$ (map of branch point $w=0$)

F_2 $(=1/\sqrt{2k})$ has two coincident images at $z=\infty$ (map of branch point $w=0$)

F_3 $(=0)=F_4$ has two coincident images at $z=0$ (map of branch point $w=\infty$).

The images in any curve of the system $u=k$ lie on some one curve of the orthogonal system $v=c$. The two systems have common foci,

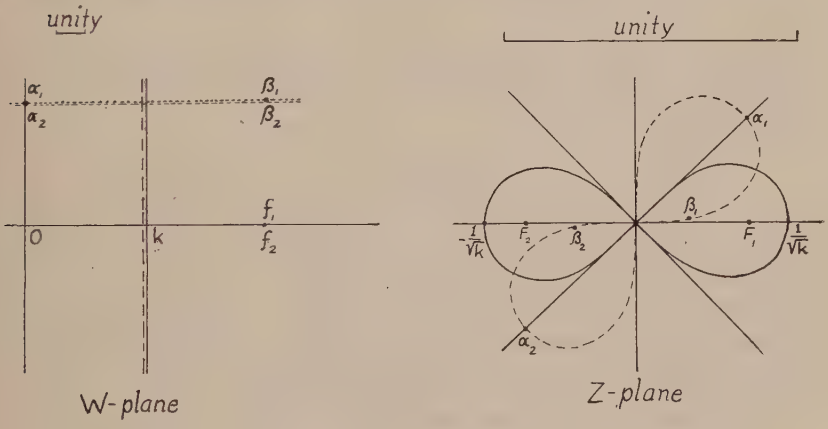


Fig. 3

F_3 and F_4. No two curves of either (or both systems) have the same F_1, F_2 foci. From the equation of images $c\bar{c}x^2\bar{y}^2=x^2+\bar{y}^2$ we find the foci of the lemniscate

$$c\bar{c}x^2\bar{x}^2=x^2+\bar{x}^2$$

[1] Functions of $\bar{z}$ are equivalently functions of z. The above function $1/\bar{z}^2$ can be written $w=\frac{1}{\bar{z}^2}=\frac{z^2}{(z\bar{z})^2}=\frac{z^2}{|z|^4}$. This particular mapping equation does not reverse angles, since w is a function of the *reciprocal* of the *conjugate* of z.

to be 0, 0, $+1/|c|$, $-1/|c|$ by equating to zero the discriminant of the quadratic in $\bar{y}$.

A SELF-INVERSE CARTESIAN

Weierstrass' p-function defined for a rectangular lattice (ω_1, and ω_2/i real) has invariants g_2 and g_3 which are real. If g_2 and g_3 are real, the coefficients in the expansion of p (being rational functions of g_2 and g_3) are real. Now any function with real coefficients is transformed into the *conjugate function* if its argument is replaced by the *conjugate argument*. Hence for a rectangular lattice $p(\bar{u}) = \bar{p}(u)$, if $u = \alpha + i\beta$, $\bar{u} = \alpha - i\beta$, $u + \bar{u} = 2\alpha$. Writing $x = p(u)$, $\bar{x} = p(\bar{u})$, $a = p(2\alpha)$, a constant, if α is held fast; as β varies, u runs on a line. From the addition theorem for $p(u+\bar{u})$,

$$4\left(x+\bar{x}+a\right)\left(ax\bar{x}-\frac{g_3}{4}\right)-\left[x\bar{x}+a(x+\bar{x})+\frac{g_2}{4}\right]^2=0\,.$$

This self-conjugate equation ($a = p[2\alpha]$ is real) represents a Cartesian curve. Hence $x = pu$ maps lines onto cartesians. Write the "equation of images"

$$4\left(x+\bar{y}+a\right)\left(ax\bar{y}-\frac{g_3}{4}\right)-\left[x\bar{y}+a(x+\bar{y})+\frac{g_2}{4}\right]^2=0\,.$$

For any determinate x this is a quadratic in $\bar{y}$. Hence every x has two images, y_1 and y_2. Each of these points has two images, and so on. Whether this system of images multiplies indefinitely or forms a finite closed system can be determined thus:

First find the foci. Two images y_1 and y_2 of any point x will coincide (determining x as a focus) if the discriminant of the quadratic in $\bar{y}$ vanishes. The equation of condition is

$$(4a^3-g_2a-g_3)(4x^3-g_2x-g_3)=0\,.$$

The roots of this equation in x (assumed to be a quartic) are:

$$\infty \equiv p(0)\,, \qquad e_1 \equiv p(\omega_1)\,, \qquad e_2 \equiv p(\omega_2)\,, \qquad e_3 \equiv p(\omega_3)\,.$$

The foci are consequently finite in number. Next, to facilitate the

finding of the image points, find the circle in which the two branches (ovals) are inverse; do likewise for the orthogonal system. In the equation

$$4\left(x+\bar{x}+a\right)\left(ax\bar{x}-\frac{g_3}{4}\right)-\left[x\bar{x}+a(x+\bar{x})+\frac{g_2}{4}\right]^2=0$$

if

$$x\equiv p(u)\equiv p(\alpha+i\beta) \text{ and } 2\alpha=\omega_1$$

then

$$a\equiv p(2\alpha)=p(\omega_1)=e_1 .$$

Now

$$\frac{g_2}{4}=-\sum e_1e_2 \quad \text{and} \quad \frac{g_3}{4}=e_1e_2e_3$$

hence

$$[x\bar{x}+e_1(x+\bar{x})-\Sigma e_1e_2]^2=4(x+\bar{x}+e_1)(e_1x\bar{x}-e_1e_2e_3)$$

which reduces to

$$(x-e_1)(\bar{x}-e_1)=(e_2-e_1)(e_3-e_1)$$

a circle, center e_1 and radius $\sqrt{(e_2-e_1)\ (e_3-e_1)}$.

If α is a variable and β is held fast, we get

$$4\left(x+\bar{x}+b\right)\left(bx\bar{x}-\frac{g_3}{4}\right)-\left[x\bar{x}+b(x+\bar{x})+\frac{g_2}{4}\right]^2=0$$

where $b\equiv p(2\beta)$. If $2\beta=\omega_2$, then

$$p(2\beta)=p(\omega_2)=e_2 ,$$

whence, as above,

$$(x-e_2)(\bar{x}-e_2)=(e_2-e_1)(e_2-e_3)$$

a circle, center e_2 and radius $\sqrt{(e_2-e_1)(e_2-e_3)}$. Hence the quarter-period lines $\alpha=2\omega_1/4$, $\beta=2\omega_2/4$ map into circles. The solid lines $\alpha=2\omega_1/8$ and $\alpha=14\omega_1/8$ map into the outer oval; $\alpha=6\omega_1/8$ and $\alpha=10\omega_1/8$ into the inner oval. In Figure 4, the parent-point u which generates the system of images is put on the quarter-period lines.

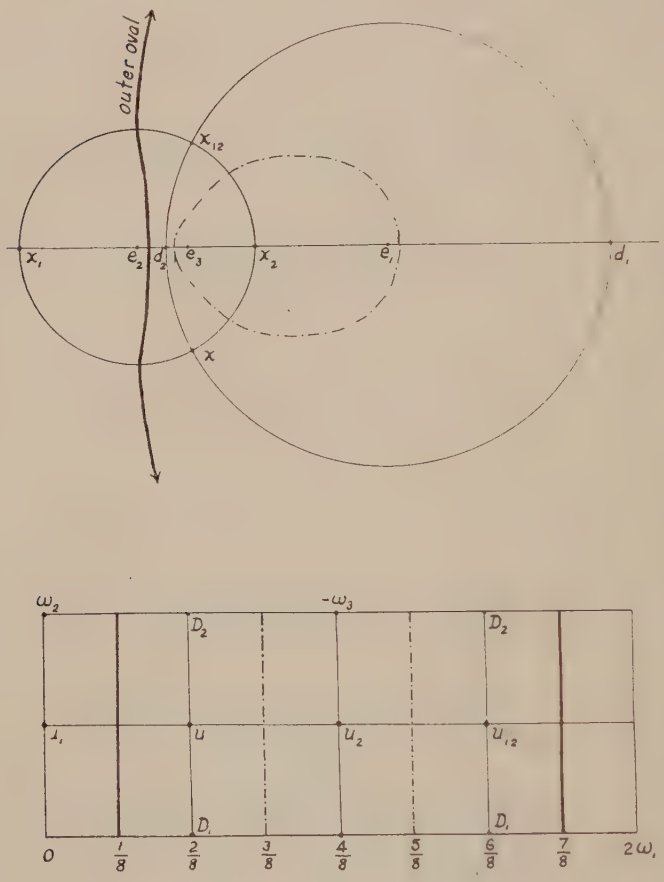


Fig. 4

The images of u and the images of images *ad infinitum*, in the four lines 1/8, 7/8, 3/8, 5/8 of $2\omega_1$, are four and only four; viz., u_1, u, u_2, u_{12} (mod $2m\omega_1$). If u runs to 1/8 of $2\omega_1$, so does u_1, while u_2, u_{12} run to 5/8 of $2\omega_1$. Hence if the parent-point x is on the Cartesian, all image points are on the Cartesian. Hence the function $p(u)$, defined for a rectangular lattice, maps the 1/8, 7/8, 3/8, 5/8, of $2\omega_1$ lines into a self-inverse Cartesian. Similarly for the orthogonal Cartesian: α variable and $\beta = 1/8$ etc. of $2\omega_2$. It is obvious that a point and its images in any Cartesian lie on an orthogonal Cartesian. In

the figure as drawn, the circle (center e_2) is a limiting case (the two branches coincident) of the orthogonal Cartesians.

The images of ∞ and e_1 are coincident at d_1

The images of e_2 and e_3 are coincident at d_2.

d_1 and d_2 are the Jacobian or double points. If one of these (say d_2) is on the circle of an inversion and the other is sent to ∞, the circle on diameter $d_1\ d_2$ is sent into a line and the two ovals are inverted into Cassinians which are images in the line. It is possible to map self-inverse Cassinians onto a line of a two-sheeted Riemann surface (one line in each sheet). The functional relation is

$$w=\frac{1}{z^2-a^2}$$

and also onto the rectangular hyperbola by the relation

$$w^2=\frac{1}{z^2-a^2}$$

where a is any constant.

CARTESIANS NOT SELF-INVERSE

For all values of a differing from those discussed above, the Cartesians are not self-inverse. However, if a is an aliquot part of a period, the number of images (in any period parallelogram) is finite, and consequently the map in the x-plane will be a Cartesian in which the number of images of a point (and the images of images) is finite. By dividing the period parallelogram by other properly chosen vertical lines, each individual image (u-plane) can be confined to its own "strip." These vertical lines will map (by fours) into Cartesians, including the limiting cases of the circle (map of $1/4$, $3/4$ period) and the line segments e_1, e_3, and $e_2\ \infty$ (maps of the $0/2$, $1/2$ period). The x-plane is divided by these Cartesians in such a way that if the region bounded by any two consecutive curves is "reflected" in the boundary, and the process repeated, the whole x-plane is covered once and only once at every point. This is analogous to the "triangulation" of the plane by reflecting a $30°$-$60°$-$90°$ triangle in its sides. If a function is defined in the region bounded

by any two above-mentioned consecutive Cartesians and takes real values on the boundary, it can be continued analytically by Schwarz's mirror method.

ELLIPSES AND HYPERBOLAS

If we use the functional relation

$$2z = x + \frac{1}{x}$$

and map radial lines and concentric circles from the x-plane to the z-surface, there arises a system of confocal hyperbolas and ellipses. The branch points are $+1$ and -1, and are the foci. Two rays equally inclined to the real axis in the x-plane map into the same hyperbola. The base circle maps into the line segment $+1$ to -1. Any two circles whose radii are reciprocals map into the same ellipse (described in opposite senses), in the upper and lower sheet.

Assume a point in the x-plane and find its images in the radial lines and then the images of images and so on. In general the images will multiply indefinitely. But if the angle of the ray with the real axis is an aliquot part of π, the images will be finite in number. If x and its first image run together, all will run together in pairs. If the angle is π/n, where n is rational, the double images will be found at the points

$$\pi/n, \quad -3\pi/n, \quad 5\pi/n, \quad -7\pi/n \ldots . \quad (-1)^{n-1}(2n-1)\pi/n .$$

Let $t^n - 1 = 0$. t_i is one of the n, nth roots of unity. If a pair of images run together on the ray $rt_i^{\frac{1}{2}}$, the position of the other images will be on the rays

$$rt_1^{\frac{1}{2}}, \quad rt_2t_1^{\frac{1}{2}}, \quad rt_3t_1^{\frac{1}{2}}, \ldots .$$

or on the rays

$$rt_1^{\frac{1}{2}}, \quad rt_1^{\frac{3}{2}}, \quad rt_1^{\frac{5}{2}}, \ldots .$$

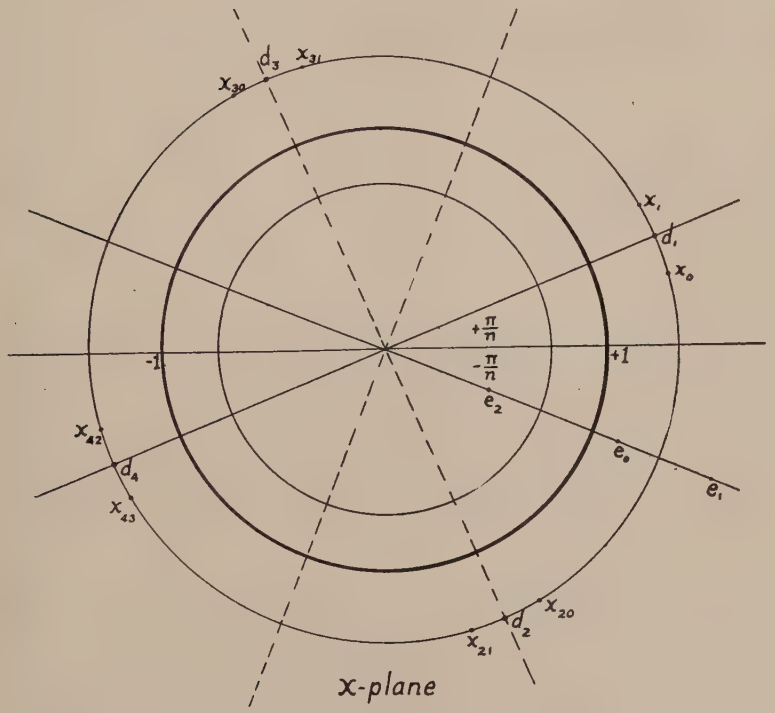

d_3
x_{31}
x_{30}
x_1
d_1
x_0
$+\frac{\pi}{n}$
$-\frac{\pi}{n}$
-1
+1
e_2
x_{42}
d_4
x_{43}
e_0
e_1
x_{20}
d_2
x_{21}
X-plane

Fig. 5a

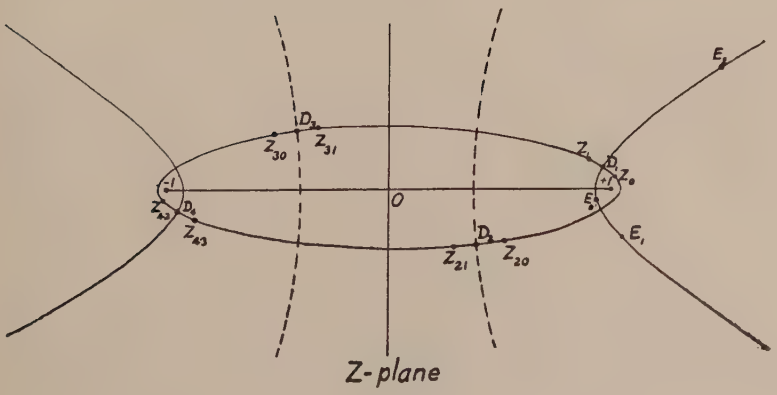

E_2
D_3
Z_{30}
Z_{31}
Z_1
D_1
Z_0
-1
+1
O
Z_{42}
D_4
Z_{43}
E_0
E_1
D_2
Z_{21}
Z_{20}
Z-plane

Fig. 5b

If we allow r both plus and minus values, and use the map equation $2z=x+1/x$, the result is the system of hyperbolas

$$2z=\left(rt_1^{\frac{1}{2}}+\frac{1}{rt_1^{\frac{1}{2}}}\right); \qquad 2z=\left(rt_1^{\frac{2}{2}}+\frac{1}{rt_1^{\frac{2}{2}}}\right); \text{ etc.}$$

Form the product.

$$\prod_{m=1}^{n}\left(rt_1^{(2m-1)/2}+\frac{1}{rt_1^{(2m-1)/2}}-2z\right)=0\,.$$

It is an equation of degree $2n$ in r and is reducible, and represents a curve of degree $2n$ also reducible. It is evidently self-inverse. In the x-plane the images in the rays lie on a circle; hence the images in the above-mentioned curve (system of hyperbolas) lie on an ellipse. Their position is determined by the intersection with the ellipse of a line parallel to the tangent to the ellipse where it crosses the hyperbola.

In the x-plane the point $+1$ has images in the rays through π/n and $-\pi/n$ at the points $2\pi/n$ and $-2\pi/n$ on the base circle. Both these points map onto the same point in the z-plane, while $+1$ (x-plane) maps onto $+1$(z-plane). This point is, then, the focus of all hyperbolas. Likewise the images of $+1$ in two circles whose radii are reciprocal map into the same point; hence, $+1$ is the focus of all ellipses. Similarly, -1 is also a focus of all hyperbolas and ellipses. Assuming two circles of reciprocal radii (which map into the same ellipse), the images therein never close in, but march to infinity and zero (x-plane) along the same ray. In the z-plane, the images of images march to infinity along a confocal hyperbola. Hence no ellipse (excluding the circle) is a self-inverse curve. As in the case of the Cartesians that are not self-inverse, if the parent-point is confined to the region bounded by two rays whose angle is an aliquot part of π, or the ring between two circles (of reciprocal radii, which are rational multiples of unity), the image points are never confounded. In the case of the ellipse, however, the image points are dense at ∞. In the case of the hyperbola, the plane is covered once by reflecting a region bounded by two consecutive curves, over the boundary. This, again, is the analogue of the "triangulation" of the plane, as explained above.

IMAGE POINTS AND GREEN'S FUNCTION

THE POTENTIAL

If we consider the attraction between a particle M and a number of particles P_i, the force of attraction is a function $f_i(r_i)$ of the distance r_i between M and the participles. The sum of the components along the axes, of all the attractions between M and P_i, is

$$X_1=\Sigma f_i'(r_i)(a_i-x_1)/r_i$$
$$X_2=\Sigma f_i'(r_i)(b_i-x_2)/r_i$$
$$X_3=\Sigma f_i'(r_i)(c_i-x_3)/r_i$$

where a_i, b_i, c_i, are co-ordinates of P_i; x_1, x_2, x_3 are co-ordinates of M; and $r_i^2=(a_i-x_1)^2+(b_i-x_2)^2+(c_i-x_3)^2$. X_1, X_2, X_3 . . . . are evidently the partial derivatives $\delta/\delta x_1$, $\delta/\delta x_2$, $\delta/\delta x_3$, ... of one and the same function V, where $V=-\Sigma f_i(r_i)$. This function is called, in a generic sense, the potential.

If the law of attraction is the same between M and all the particles P_i, then f_i is the same for all values of i. As a further restriction, if the space is three dimensional and the law of attraction is the inverse-square law, the potential is called Newtonian. In this case, if the mass at M is unity and m_i at P_i, then

$$X=\Sigma m_i(a_i-x_1)/r_i^3\ ,\quad Y=\Sigma m_i(b_i-x_2)/r_i^3\ ,\quad Z=\Sigma m_i(c_i-x_3)/r_i^3\ ,$$

and

$$V=\Sigma m_i/r_i.$$

In certain distributions of the attracting matter, e.g., an infinitely long thin rod, the force varies inversely as the distance from the rod, although the individual particles attract as the "inverse square." In this case

$$f_i'(r_i)=m_i/r_i$$

whence

$$f(r_i)=m_i\log\frac{r_i}{r_0}\ ;\quad r_0\ \text{constant}\ .$$

$$V=-\Sigma m_i \log \frac{r_i}{r_0}, \quad X=\frac{\delta V}{\delta x}=\Sigma m_i(a_i-x)r_i^2, \quad Y=\frac{\delta V}{\delta y}=\Sigma m_i(b_i-x)/r_i^2, \text{ etc.}$$

Here V differs from the Newtonian potential. Although finite and continuous with its first and second derivatives at a finite distance outside the mass, it becomes at ∞ infinite like $-\log 1/r$, and within the mass ∞ like $\log 1/r$. It has the essential characteristic of the potential function, viz., that the force components along the axes are the partial derivatives of one of the same function—this function V. It satisfies Laplace's equation $\Delta_2 V=0$ only in the two-dimensional form. Although the term "logarithmic potential" is used for the function V in any dimension, it is generally restricted to this function in the plane. In this restricted sense V is one of the conjugate functions of a function of a complex variable.

GREEN'S FUNCTION

Green's function belongs to the category of potential functions. The general problem in the theory of potential—Dirichlet's problem—is to find the potential at any point when the value either of the potential itself or its normal derivative is prescribed on the boundary of a definite region. In general, both cannot be prescribed arbitrarily. For example, a uniformly heated conducting sphere may have its surface kept at the constant temperature zero by immersion in ice water, or the normal flow through the surface may be reduced to zero by an insulating jacket. But both conditions cannot be assumed to exist simultaneously unless the whole sphere is at zero degrees, in which case there is no problem to be solved. If one or other boundary condition is prescribed, the procedure is as follows: Assume Green's theorem in the form

$$\int\left(U\Delta V-V\Delta U\right)dt=\int\left(U\frac{\delta V}{\delta n}-V\frac{\delta U}{\delta n}\right)ds. \qquad (1)$$

In two dimensions, the first integral is taken over the area bounded by the curve, and the second the line integral along the curve. Since U and V are any functions whatever, we can assume that V satisfies Laplace's equation throughout the whole plane, and that U has the singularities of the logarithmic potential; i.e., $U=\log 1/r$. If a circle of small radius be drawn about the singular point, then both

ΔU and ΔV are zero in the region between the circle and the curve C. Formula (1) then becomes

$$\int\left(\log 1/r\frac{\delta V}{\delta n}-V\frac{\delta \log 1/r}{\delta n}\right)ds=0 \qquad (2)$$

where the integral is taken over the complete boundary, the circle and C. The integral (2) is

$$\int_c\left(\log 1/r\frac{\delta V}{\delta n}-V\frac{\delta \log 1/r}{\delta n}\right)ds+\int_0\left(\log 1/r\frac{\delta V}{\delta n}-V\frac{\delta \log 1/r}{\delta n}\right)ds . \qquad (3)$$

The second of these reduces to $-2\pi V$. Hence

$$V=\frac{1}{2\pi}\int_c\left(\log 1/r\frac{\delta V}{\delta n}-V\frac{\delta \log 1/r}{\delta n}\right)ds . \qquad (4)$$

This gives V for any point in the plane in terms of $\delta V/\delta n$ and V on the boundary. But, as remarked above, both cannot be taken arbitrarily; hence one must be eliminated. This can be accomplished by means of a special function which has the following properties: (1) It vanishes on C. (2) At some point m it becomes infinite like $\log 1/r$. (3) At ever other point within C it is regular and satisfies Laplace's equation. This function is called the Green function and is denoted by $G(x, y, a, b)$. If in equation (1) instead of choosing $U=\log 1/r$ we make $U=G-\log 1/r$, equation (1) becomes

$$0=\int_c\left[\left(G-\log 1/r\right)\frac{\delta V}{\delta n}-V\frac{\delta}{\delta n}\left(G-\log 1/r\right)\right]ds$$

$$=\int_c\left(G\frac{\delta V}{\delta n}-V\frac{\delta G}{\delta n}\right)ds-\int_c\left(\log 1/r\frac{\delta V}{\delta n}-V\frac{\delta}{\delta n}\log 1/r\right)ds$$

but G vanishes on C and the second integral equals $2\pi V$. Hence

$$V_p=-\frac{1}{2\pi}\int_c V_c\left(\frac{\delta G}{\delta n}\right)_c ds .$$

If, then, the potential V_p for any point p is required when V_c, the value of the potential, is given on the boundary, the problem can be solved if G is known for the contour.

The connecting link between potential theory and function theory is precisely the Green function; since it is G that maps the contour on a line and e^G that maps it onto a circle. If G can be shown to exist for every contour, Riemann's theorem is established. If G is found for various contours, the function that maps them onto the circle is known. Hence the importance of finding the Green function.

The "method of images" is a convenient tool in the construction of the Green function for a contour in a number of cases.

If $f(x, \bar{x}) = 0$ is self-conjugate, we have seen that it represents a curve C_f, and $f(x\bar{y}) = 0$ determines the images $\bar{y}_i$ in this curve, of any point x, and conversely. Since f is real along C_f (i.e., $f = \bar{f}$ on C_f) and is defined for points x not on C_f, it can be continued analytically over C_f (if known on one side) by associating with f at the points x the conjugate $\bar{f}$ at the image points y. If now the plane of C_f is mapped by the relation $g = g(x)$ onto the plane of C_g, then when x is on C_f, g is on C_g. It has image points in C_g, viz., $\bar{g}(\bar{y})$, the correspondents of the image points of x in C_f. The function $\bar{g}$ is the analytic continuation of g across C_g. If now g is such a function that $2G = g + \bar{g}$, where G is the Green function, then, since by definition the Green function is zero on the contour,

$$g(x) + \bar{g}(\bar{y}) = 0$$

defining C_g as the imaginary axis, which in turn can be mapped onto the unit circle. The Green function has a logarithmic infinity at some point, say x_0. Then in the vicinity of x_0

$$g(x) = \log \frac{1}{x - x_0} + R(x)$$

where R is regular at x_0. The infinity of g, viz., x_0, will have images in C_f. Call any one $\bar{y}_0$. Then from the relation $g(x) + \bar{g}(\bar{y}) = 0$

$$g(y) = -\bar{g}(\bar{x}) = -\log \frac{1}{\bar{x} - \bar{x}_0} + \bar{R} .$$

Since corresponding image points are solutions of $f(x\bar{y})=0$, the equation $\bar{x}(y)-x_0=0$ will have as solutions $y=y_0$, where y_0 is an image of $\bar{x}_0$. Whence

$$g(\bar{y})=-\log\frac{1}{\bar{y}-y_0}.$$

The function g, therefore, which at x_0 becomes infinite like $+\log\frac{1}{x-x_0}$, at y_0 (an image of x_0) becomes infinite like $-\log\frac{1}{y-y_0}$. The function e^g has therefore a simple pole at x_0 and a simple zero at the image points y_0. Hence the theorem: Corresponding image points are simple poles and zeros (respectively) of e^g.

This theorem opens the way for the construction of the Green function G for various contours, by determining e^g for the regions bounded by the contours. From the discussion of image points given above, it is seen that when various regions of the plane are marked off and bounded by various contours, and all the images of a point and images of images in the contour are constructed, then (A) some regions contain one only of these points, and (B) others more. Of regions of the latter class, some (B) *a*) have the images congruent modulo $2m\omega$, or mod $2m\omega+2n\omega'$ (cf. self-inverse Cartesian); (B) *b*) others not so. From these cases it was shown how in some instances the plane can be "triangulated," in others not. In case (A), e^g is a uniform function; in case (B) *a*), also uniform, but singly or doubly periodic. In case (B) *b*), e^g is not uniform. In any of the above cases, if the crop of images is "simply finite," e^g is algebraic (uniform or multiform). If finite modulo $2m\omega$ or $2m\omega+2n\omega'$, e^g is a (uniform or multiform) algebraic function of singly or doubly periodic functions. If there is an infinite number of images not congruent to a finite set, then e^g is in no sense algebraic, nor uniform (cf. Cartesians in which α is *not* an aliquot part of a period).

These assertions seem true as deductions from the above general discussion and the theorem on the zeros and poles of e^g.

II. FUNCTION-THEORY PROOF OF RIEMANN'S THEOREM

SUMMARY OF METHOD: THE FORMULATION OF THE MAPPING PROBLEM

The purpose is to prove the existence of a function that will map a region conformally on the unit circle. The first region that comes to our notice is the absolutely boundless region (or a region with no boundary points). Such a region cannot be mapped onto the unit circle. This follows from the fact that a function which is everywhere regular must be a constant, and consequently no mapping will result. But a region with only a single point as boundary must likewise be excluded. For by a linear transformation the point can be sent to infinity, and a function which is regular for all finite values of the argument and whose absolute value lies below a fixed upper bound is also a constant. (The mapping function must have an upper bound of its absolute value, since the region is to be mapped onto the inner points of the unit circle.) The proof is given below, Lemma II.

Accordingly we consider a region having *at least* two boundary points. This region, accordingly, will not be, in general, simply connected. Make it so by a cross-cut. The region does not contain any outer points, but only inner and boundary points. If the original boundary points are a and b, these can be sent to zero and infinity by the transformation $z'=\frac{z-a}{z-b}$. If this region is transformed by $z''=1: z'$, the region B' will go over into B''. In the z'' plane there are outer, boundary, and inner points. Now B'' can be transformed into a region which lies *entirely within* the unit circle. We can therefore without loss of generality restrict our problem to the proof of the existence of a function which will map onto the unit circle a region lying entirely within the unit circle. If the region be such that there are initially outer points, and if z_0 be one of these, then the first operation consists in applying the transformation $z'=\frac{\alpha}{z-z_0}$ (where α is a sufficiently small quantity). Since z cannot equal z_0 (z is an inner point), the original region can be made to lie entirely within the unit circle.

The distinction (first made by Schwarz) between the mapping of inner points onto inner points, and boundary points onto boundary points, of the circle disentangles to some extent the snarls of an inherently difficult proof. The mapping of inner points is possible sometimes when the mapping of boundary points is not.[1]

The outline of the method of proof due to Koebe[2] is as follows: Suppose an endless series of one-to-one conformal mapping operations. The first region B_1 passes over into a region B_2; then B_2 into B_3 and so on indefinitely. The mapping function is of such a character that all the regions B_n are in the interior of the unit circle. If z_1 is the variable point in the region B_1, then z_2, z_3, z_4 . . . . are the maps of z_1, in B_2, B_3. . . . . If $1-\epsilon_n$ is the radius of the largest circle C_n (center O) which lies completely in B_n, and ϵ_n is the width of the ring r_n between the unit circle and C_n, then $\epsilon_1>\epsilon_2>\epsilon_3$ . . . . , and $\lim_{n\to\infty} \epsilon_n=0$; that is, the regions tend continuously to fill out the unit circle. The contours are "nested," approaching closer and closer to the periphery of the circle.

The region B_{n+1} is derived from the region B_n thus: On the boundary L_n pick out the point a_n that lies closest to O. Spread B_n on a two-sheeted surface that has a branch point at a_n and the unit circle (doubly described) as its boundary. Call this surface W_n.

The nth "nesting" operation by which the region B_n goes into the region B_{n+1} is now defined as that mapping operation by which the surface W_n is mapped conformally one-to-one, on the "flush" surface of the unit circle. This can be accomplished by first mapping the unit circle onto itself so that the point a_n goes into O (a turning to bring a_n to the real axis, followed by a real linear transformation). By this transformation W_n goes into W'_n with O as the winding-point. Next, by a square-root operation W'_n is mapped onto the flush unit circle with O remaining fixed. Any direction R' can by a simple turning be made to coincide with a prescribed direction R, thus making the mapping unique. The mapping function $z_{n+1}\equiv \phi_n(z_n)$ is a function compounded of elementary functions. The in-

[1] Cf. Carathéodory, *Math. Ann.*, LXXII (1912), 107.

[2] *Crelle*, Vol. CXLV.

verse of ϕ_n is a rational function of the second degree, $z_n = \psi_n(z_{n+1})$, which vanishes at O and some other point in the interior and nowhere externally to the unit circle. On the unit circle its absolute value is unity. The quotient z_n/z_{n+1} is a rational function of z which takes the maximum of its absolute value on the unit circle. Hence $|z_n/z_{n+1}| < 1$ for all inner points of the unit circle. The function $\phi_n(z)$ therefore maps B_n onto a region B_{n+1} in such way that all points of B_{n+1} lie farther from O than corresponding points of B_n (except the point O itself and such points of the boundary of B_n as perchance lie on the unit circle).

PRELIMINARY LEMMAS

LEMMA I[1]

"An analytic function $f(z)$ which differs from a mere constant cannot have at a regular point z_0 (not on its boundary) a maximum of its absolute value, nor a minimum (unless it be zero at z_0), nor a maximum or minimum of the real part of $f(z)$."

Write:

$$f(z) = f(z_0) + a_n(z - z_0)^n + (z - z_0)^{n+1} f_1(z)$$

where $a_n \neq 0$, $n = 1, 2, 3 \ldots$, and $f_1(z)$ is regular at z_0. In the neighborhood of z_0 sufficiently small, $a_n(z - z_0)^n$ is preponderant. In this region draw a vector $z - z_0$ in the same sense as $f(z_0)$, and the first statement is evident; reverse the vector, and the second statement is seen to be true; point the vector parallel to the real axis, then reverse it, and the third and fourth statements are evident.[2]

LEMMA II[3]

"A function which has a finite upper limit of its absolute value and is also regular for all finite values of the argument is a constant."

---

[1] Koebe, *Crelle*, CXLV, 179.

[2] There is an analogous lemma in potential theory; cf. Poincaré, *Potentiel Newtonien*, p. 136.

[3] Koebe, *Ibid.*, p. 181.

Thus, let $F(z)$ be such a function and let z_0 be a regular point of $F(z)$. The function

$$\phi(z)=\frac{F(z)-F(z_0)}{z-z_0}$$

is for every finite z regular and is indefinitely small, for z indefinitely large (since $F(z)$ has a finite upper bound). Hence if $\phi(z)$ did not vanish identically, it would have a maximum of its absolute value for some finite value of z, hence not on its boundary (since a finite z does not yield a boundary point). Therefore $\phi(z)$ vanishes identically. Therefore

$$F(z)\equiv F(z_0)=\text{a constant.}$$

LEMMA III[1]

The following is Carathédory's proof:

"If $f(z)$ is analytic; regular for $|z|<1$; and if $|f(z)|\leq 1$ and $f(0)=0$, then $|f(z)|\leq|z|$."

Proof: $\phi(z)\equiv\frac{f(z)}{z}$ is regular for $|z|<1$; assume $r<1$. The maximum of $|\phi(z)|$ within the circle $|z|\leq r$ is reached on the periphery $|z|=r$. On this periphery

$$\left|\phi(z)\right| \equiv \left|\frac{f(z)}{z}\right| = \left|\frac{f(z)}{r}\right| \leq \frac{1}{r}$$

since $|f(z)|\leq 1$ by supposition. Whence $\left|\frac{f(z)}{z}\right|\leq\frac{1}{r}$ or $|f(z)|\leq\frac{|z|}{r}$. Since this is true for any $r<1$, $|f(z)|\leq|z|$.

LEMMA IV[2]

"If $w=u+iv$, $z=x+iy$, and if the circle $|z|\leq a$ is mapped onto another region by the non-linear relation:

---

[1] Schwarz's lemma, *Gesamm. Abh.*, Vol II, enunciated and proved in various ways by Bieberbach, *Korforme Abbild.*, p. 33; Carathéodory, *Math. Ann.*, LXXII, 110; Koebe, *op. cit.*, p. 190.

[2] Bieberbach, *Rend. del Circ. Mat. Di Palermo*, XXXVIII (1914), 105; and *Konforme Abbild.*, p. 95.

$$w=f(z)=d_1z+d_2z^2+d_3z^3+ \ . \ . \ . \ . \ .$$

The area A of the mapped region is greater than $\pi d_1 \bar{d}_1 a^2$."

Proof: $A=\int\int du\ dv$. By Duhamel's theorem, this can be written $A=\int\int |i\ ds\cdot ds|$.

$$\therefore A=\int\int (du^2+dv^2)=\int\int dw\cdot d\bar{w}=\int\int f'(z)\bar{f}'(z)dz\ d\bar{z}$$

$$A=\int\int f'(z)\bar{f}'(\bar{z})(dr^2+r^2\ d\phi^2)=\int\int f'(z)\bar{f}'(\bar{z})r\ dr\ d\phi$$

writing $z=re^{i\phi}$, and applying Duhamel's theorem again. Inserting the values of $f'(z)$ and $\bar{f}'(\bar{z})$,

$$A=\int_0^a r\ dr^2\int_0^{2\pi} d\phi[d_1+2d_2re^{i\phi}+3d_3r^2e^{2i\phi}+ \ . \ . \ . \ . \ ]$$
$$[\bar{d}_1+2\bar{d}_2re^{-i\phi}+ \ . \ . \ . \ . \ ] \ .$$

The last written product will yield terms independent of ϕ and the rest of the form $Ce^{ki\phi}$ where C does not involve ϕ. The latter all vanish in the integration. Hence

$$A=2\pi[d_1\bar{d}_1a^2/2+4d_2\bar{d}_2a^4/4+9d_3\bar{d}_3a^6/6+ \ . \ . \ . \ . \ ]>\pi d_1\bar{d}_1a^2 \ . \quad \text{Q.E.D.}$$

If A differs from πa^2 by ϵ, i.e., $|A-\pi a^2|=\epsilon$, then

$$|\ (d_1\bar{d}_1-1)\pi a^2+4\pi d_2\bar{d}_2a^4/2+ \ . \ . \ . \ . \ |=\epsilon.$$

Place $|d_1\bar{d}_1-1|=\delta$; then

$$2\pi(4d_2\bar{d}_2a^4/4+9d_3\bar{d}_3a^6/6+ \ . \ . \ . \ .)<\epsilon+\delta\pi a^2$$

since all the terms on the left are positive. Hence also

$$n^2d_n\bar{d}_na^{2n}/2n<\epsilon+\delta\pi a^2 \text{ or } d_n\bar{d}_n<\frac{2}{n}(\epsilon+\delta\pi a^2) \ .$$

From this we find an upper limit for $|f(z)-z|$. If $|z|\leq\rho a$, $\rho<1$,

$$|f(z)-z|=|(d_1-1)z+d_2z^2+d_3z^3+ \ . \ . \ . \ . \ |\leq|(d_1-1)z|+|d_2z^2|+ \ . \ . \ . \ .$$
$$\leq|d_1-1|\rho a+|d_2|\rho^2a^2+|d_3|\rho^3a^3+ \ . \ . \ . \ .$$
$$\leq|d_1-1|\rho a+|d_m|a^m\frac{\rho^2}{1-\rho}$$

where $|d_m|a^m \geq |d_n|a^n$, $n = 2, 3, 4, \ldots$. Now

$$|d_m|a^m = \sqrt{d_m \bar{d}_m a^m} = \sqrt{d_m \bar{d}_m a^{2m}} < \sqrt{\frac{2}{m}(\epsilon + \delta\pi a^2)} .$$

If $|d_1 - 1| \leq \delta$, then

$$|f(z) - z| < \delta\rho a + \sqrt{\frac{2}{m}(\epsilon + \delta\pi a^2)}$$

which vanishes with ϵ and δ.

Cor: Let $f_s(z)$ be a sequence of functions regular in the circle $|z| < a$, and let $f_s(0) = 0$ and $\operatorname{Lim}_{s\to\infty} f_s'(0) = 1$ and $\operatorname{Lim}_{s\to\infty} A_s = \pi a^2$; then $\lim_{s\to\infty} f_s(z) = z$, since under these conditions ϵ and δ both vanish.

Definition: "flush," "equivalently flush," "non-flush."—Regions may be divided into two classes, those in which (A) no part overlies; (B) some part overlies another part. The latter may be subdivided: those in which a continuous path (e.g., a polygon, or simple curve without double points) can be drawn which (*a*) divides the region

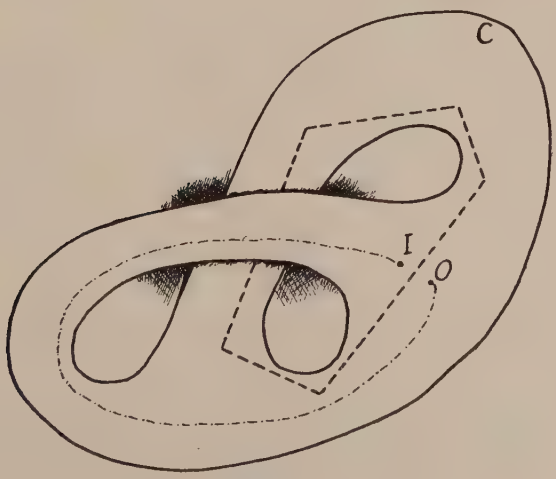


Fig. 6

into two parts; (*b*) does not divide the region into two parts. (A) are flush regions; (B) *a*) equivalently flush; (B) *b*) non-flush.

A sample of a non-flush region is illustrated in Figure 6 unshaded. (It is simply connected: its boundary C forms a continuous point set.) It is possible to go, by a continuous path, from the inside I to the outside O of the polygon without crossing the boundary lines of the polygon.

PROOF OF RIEMANN'S THEOREM

CONSTRUCTION OF FUNCTION, ACCORDING TO KOEBE'S DESCRIPTION

The initial region is supposed to have been transformed into a region B which lies entirely within the flush surface of the unit circle, and contains the origin. Mark the point a (on the contour) nearest the origin. Lay the region on a two-sheeted surface and make a a branch point; draw the branch cut outward across the unit circle. Describe the unit circle in both sheets (a continuous curve). Assume two complex quantities w and z (on the same surface), functions one of the other. They must both be zero together. Also if $|w| = 1$, then $|z| = 1$ (and conversely); the unit circle always goes into itself. The following functional relation

$$\frac{\sqrt{a} - w}{1 - \sqrt{\bar{a}} w} = \sqrt{\frac{a - z}{1 - \bar{a} z}}$$

has these characteristics. Or, solved for z:

$$z = w \frac{A - w}{1 - A w}$$

where $A = \dfrac{2\sqrt{a}}{1 + \sqrt{a\bar{a}}}$. z is a rational function of w of the second degree

which vanishes for $w = 0$, and at one other point $w = A$ within the unit circle. But if w is restricted to the region B, then w cannot equal A. Rotate the plane so that a is real, then A is also real. Moreover,

$A > a$ for $o < a < 1$. Since for w real, $w < a$ in B ($w \neq A$ in B). Hence the origin is not covered twice; that is, the mapping is everywhere flush.

That the above functional relation maps the unit circle into itself is shown thus: If $|w| = 1$, then $\bar{w} = 1/w$. Now

$$z = w\frac{A-w}{1-\bar{A}w} = \frac{A-w}{1/w-\bar{A}} ;$$

hence for $|w| = 1$,

$$z = \frac{A-w}{\bar{w}-\bar{A}} = -\frac{A-w}{\bar{A}-\bar{w}} ;$$

hence $|z| = 1$. If then we apply Koebe's lemma that a bounded function takes the maximum of its absolute value on the boundary, it follows that $|z| = |\phi w| \leq 1$. This function then fulfills all the conditions of Schwarz's lemma. Therefore $|z| = |\phi w| \leq |w|$. Hence for the inverse function $w = f(z)$,

$$|f(z)| \geq |z| .$$

But $|f(z)| = |z|$ only on the boundary. Hence within the unit circle $|f(z)| > |z|$. Consequently mapping a region lying within the unit circle by means of $w = f(z)$, every point of the map will lie farther from the origin than its correspondent (except the origin and points which perhaps lie on the unit circle).

In the following we need $\phi'(o)$:

$$z' = \phi'(w) = \frac{A - 2\bar{w} + \bar{A}w^2}{(1-\bar{A}w)^2} ; \quad \phi'(o) = A$$

(in general not real). We wish to have $\phi'(o)$ real. Hence rewrite $\phi(w)$

$$\phi(w) = \sqrt{\frac{\bar{A}}{A}}w\frac{A-w}{1-\bar{A}w} .$$

Now $\phi'(o)=|A|$, a real positive quantity, while $|\phi(w)|$ remains unchanged, since

$$\left|\sqrt{\frac{\bar{A}}{A}}\right|=1.$$

Koebe's repeating process.—Assume the region G lying entirely within the unit circle, and containing the origin. This will map by $w=f(z)$ into a region G_1, all points farther from the origin yet within the unit circle. If a_k is as above, the point on the contour of G_k nearest the origin, then the circle $|z|\leq a_k$ lies entirely in G_k; also the circle $|z|\leq a_1$ lies entirely in G. Evidently $|a_1|>|a|$. Repeat the process: $|a_2|>|a_1|$; again, $|a_3|>|a_2|$. Hence

$$|a|<|a_1|<|a_2|<|a_3| \ldots\ldots$$

This is a monotonically increasing sequence of quantities whose absolute value cannot exceed unity. Therefore $\operatorname{Lim}_{n\to\infty}|a_n|=1$. Hence the result: By a sequence of functions

$$f_1(z);\quad f_2(z);\quad f_3(z)\ \ldots\ldots$$

built up after the manner of $f(z)$, a region G_k (inner points) can be mapped in (1, 1) correspondence, conformally on the inner points of the unit circle. In general this sequence is not finite.

Definition: "compact."—"An infinite suite of functions is said to be compact if the $\operatorname{Lim}_{n\to\infty} f_n(z)$ exists and is equal (say) to $f(z)$." A simple illustration of a compact suite, whose compactness can readily be shown by throwing the burden of proof onto a well-known convergent series, is the following:

$$z_1={}_1\phi(z)=e^{\frac{i\pi}{2}}z \qquad =\phi_1(z)$$

$$z_2={}_2\phi(z_1)=e^{\frac{i\pi}{4}}z_1=e^{i\pi\left(\frac{1}{2}+\frac{1}{4}\right)}z=\phi_2(z)$$

$$z_3={}_3\phi(z_2)=e^{\frac{i\pi}{8}}z_2=e^{i\pi\left(\frac{1}{2}+\frac{1}{4}+\frac{1}{8}\right)}z=\phi_3(z)$$

$$\lim_{n\to\infty} \phi_n(z) = -z$$

In this illustration the region G is the unit circle which is mapped onto the unit circle by a compact suite $\phi_n(z)$, whose limit $\phi(z) = -z$.

To prove the compactness of the suite $f_n(z)$, we show first the compactness of the inverse functions $z = \phi_n(w)$. We must therefore show that in every inner part region of the unit circle from a definite n (independent of m) onward and for arbitrary m

$$|\phi_{n+m}(w) - \phi_n(w)| < \epsilon$$

where ϵ is a definite positive small quantity. Write

$$\phi_{n+m}(w) \equiv \phi_n[\phi_{n,m}(w)] .$$

Now $\phi_{n,m}(w)$ is a function defined for all values of w in the whole region G_{n+m}; $\phi_{n,m}$ maps this region conformally (1, 1) on G_n; its first derivative is real and positive for $w = 0$. In order to apply corollary of Lemma IV, we show that $\lim_{n\to\infty} \phi'_{nm}(0)$ converges to unity. Write

$$\phi'_{nm}(0) = d_{nm} \ (d_{nm} \text{ real-positive}) .$$

$\phi_{nm}(w)$ is defined for every w in G_{nm} and consequently for every w in the whole circle $|w| \leq a_n$, since the latter lies wholly in G_{n+m}. $\phi_{nm}(w)$ maps this circle onto a region whose area A is greater than $\pi d_{nm}^2 a_n^2$ (cf. Lemma IV). But this same mapped region lies wholly within the unit circle (area π). Hence the inequality:

$$\pi > A > \pi d_{nm}^2 a_n^2 \quad \text{or} \quad d_{nm}^2 \leq \frac{1}{a_n^2} .$$

On the other hand, the function inverse to $\phi_{nm}(w)$ is defined for every z in the whole region G_n; hence also in $|z| \leq a_n$. This inverse function has at the origin a derivative $1/d_{nm}$ and consequently maps $|z| \leq a_n$ onto a region (area $= A'$) whose area is greater than $\pi a_n^2 / d_{nm}^2$. This area A' lies in the unit circle. Hence the inequality:

$$\pi > A' > \frac{\pi a_n^2}{d_{nm}^2}, \text{ or } d_{nm}^2 > a_n^2 .$$

Combining:

$$a_n^2 \leq d_{nm}^2 \leq 1/a_n^2 .$$

But since $\underset{n\to\infty}{\text{Lim}}\, a_n = 1$, it follows that from a given n on, the derivative $\phi'_{nm}(o)$ differs from unity by an arbitrarily small amount. We can now apply Lemma IV Cor. to $\phi_{nm}(w)$. (1) It is a function regular in the circle $|w| < |a_n|$. (2) $\phi_{nm}(o) = o$. (3) $\underset{n\to\infty}{\text{Lim}}\, \phi'_{nm}(o) = 1$. (4) Since d_{nm} and a_n both converge to the same limit unity, from a determined n on, the area of the mapped region (the map of $|w| \leq a_n$) converges to the area of the circle $|w| \leq a_n$, that is, to πa_n^2. These are the conditions in Lemma IV Cor. Therefore

$$\underset{n\to\infty}{\text{Lim}}\, \phi_{nm}(w) = w .$$

We express the limit in a form convenient for use as follows: Let B (see Fig. 7) be a region of the unit circle. Then for any arbitrary

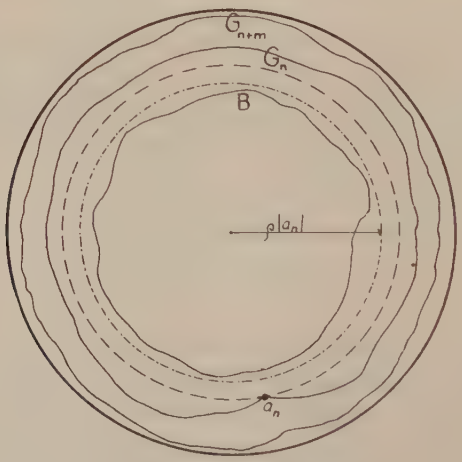


Fig. 7

ϵ an $N(\epsilon)$ can be found, so that $|\phi_{nm}(w)-w|<\epsilon$ if w lies in B and $n>N(\epsilon)$. We can always choose n so large that the region B lies entirely in G_n and in $|w|\leq\rho|a_n|$, where ρ is independent of n and $0<\rho<1$.

In the whole region B, both $\phi_{n+m}(w)$ and $\phi_n(w)$ are defined.[1] Consider now the difference

$$|\phi_{n+m}(w)-\phi_n(w)|=|\phi_n[\phi_{nm}(w)]-\phi_n(w)|$$

and put $\phi_{nm}(w)=w_1$. Then

$$|w_1-w|<\epsilon\ .$$

By Cauchy's integral formula,

$$\phi_{n+m}(w)-\phi_n(w)=\phi_n(w_1)-\phi_n(w)=\frac{1}{2\pi i}\int_{\mathfrak{I}}\phi_n(\xi)\left[\frac{1}{\xi-w_1}-\frac{1}{\xi-w}\right]d\xi$$

$$=\frac{1}{2\pi i}\int_{\mathfrak{I}}\phi_n(\xi)\left[\frac{w_1-w}{(\xi-w_1)(\xi-w)}\right]d\xi$$

$$|\phi_n(w_1)-\phi_n(w)|=\frac{1}{2\pi}\int|\phi_n(\xi)|\frac{|w_1-w|}{|\xi-w_1|\ |\xi-w|}\,|d\xi|\ .$$

Put $\xi=a_ne^{i\theta}$, $|\xi|=a_n$. $|\phi_n(\xi)|<1$, since ϕ_n is a bounded function <1. Now $|w|\leq\rho|a_n|$ and $|w_1|\leq\rho|a_n|$, while $|\xi|>\rho|a_n|$, $(\rho<1)$. Therefore $|w|<|\xi|$ and $|w_1|<|\xi|$. Also

$$|\xi-w_1|\geq|\xi|-|w_1|\geq|\xi|-\rho|a_n|\geq|a_n|(1-\rho)$$
$$|\xi-w\ |\geq|\xi|-|w\ |\geq|\xi|-\rho|a_n|\geq|a_n|(1-\rho)\ .$$

Since $\xi=a_ne^{i\theta}$; $d\xi=a_ne^{i\theta}\,d\theta$; $|d\xi|=|a_n|d\theta$. Hence:

---

[1] Since we proved $\lim_{n\to\infty}\phi_{nm}(w)=w$ for all points w in G_{n+m}, $(\phi_{nm}(w)$ is in $G_n)$, it follows that for any region B, lying entirely in the unit circle—since G_n and G_{nm} tend to fill out the unit circle as n increases—$|\phi_{nm}(w)-w|<\epsilon$ as long as w lies in B. Hence this limit does not depend on the position of w except perhaps that it must lie in B. Hence ϕ_{nm} converges uniformly to w for all $n>N(\epsilon)$. The purpose in introducing B is to insure $w\leq\rho|a_n|$.

$$|\phi_{n+m}(w)-\phi_n(w)|<\frac{1}{2\pi}\int_0^{2\pi}\frac{|w_1-w|}{|a_n|^2(1-\rho)^2}\,c_n\,d\theta$$

or

$$<\frac{|w_1-w|}{|a_n|(1-\rho)^2}<\frac{\epsilon}{|a_n|(1-\rho)^2}.$$

Hence for any assigned number $N(\epsilon\ \rho|a_n|)$ we can find a number n (by taking n sufficiently large) so that $\phi_{n+m}(w)$ converges to $\phi_n(w)$. Consequently the suite of functions $\phi_n(w)$ is compact. (Compare simple example of compact suite given above, which maps a region, the unit circle, onto the unit circle by a total rotation through π. Here

$$|\phi_{n+m}(z)-\phi_n(z)|=|ze^{i\pi\sum_1^{n+m}1/2n}-ze^{i\pi\sum_1^{n}1/2n}|$$

can be made arbitrarily small by choosing n sufficiently large.)

We must now show that the limiting function of the compact suite $\phi_n(w)$ maps the unit circle onto G.

We prove (1) that $\phi(w)$ is not a mere constant; (2) that the limiting function $\phi(w)$ maps the unit circle onto a region G' which is coincident with G wholly or in part; (3) G' is a flush region; (4) that G' is wholly coincident with G.

1. $\phi(w)$ is not a mere constant.

Proof: All functions of the suite $\phi_n(w)$ are of such character that any one of them $\phi_k(w)$ maps G_k onto G, $(k=1, 2, \ldots, n)$. Apply Lemma IV. Let $|w|\leq r$ be a circle lying wholly in G; it lies consequently entirely in every G_n. Hence all the functions $\phi_n(w)$ are defined for values of w in circle $|w|\leq r$. They map this circle onto a region (area $=A$) which by IV is greater than $\pi r^2/d_n\bar{d}_n$, where $1/d_n=\phi_n'(o)$. Again, of the suite of functions $f_n(z)$ (inverse to $\phi_n(w)$), every $f_k(z)$ is defined for all values of z in G, and consequently in the circle $|w|\leq r$, lying wholly in G. They map $|w|\leq r$ onto a region (area $=A'$) which is greater than $\pi d_n\bar{d}_n r^2$ $(=\pi|d_n|^2r^2)$, where $d_n=f_n'(o)=1/\phi_n'(o)$ and less than π. Both A and A' are less than π, since functions of both suites $\phi_n(w)$ and $f_n(z)$ map a region of the unit circle onto a

region lying entirely within the unit circle. From these inequalities we readily deduce the inequality

$$r^2 < |\phi_n'(o)|^2 < \frac{1}{r^2} .$$

Since $o < r < 1$, $\phi_n'(o)$ is not zero, and $\phi_n(w)$ is not a constant; since, moreover, r is independent of n,

$$\lim_{n \to \infty} \phi_n'(o) \equiv \phi'(o) \neq o .$$

Therefore the limiting function $\phi(w)$ of the suite $\phi_n(w)$ is not a mere constant.

2. This limiting function $\phi(w)$ maps the unit circle onto a region G' which is coincident with G, wholly or in part.

Proof: Every $\phi_k(w)$ of the suite maps inner points of G_k onto inner points of G. Proceeding to the limit, $\phi(w)$ maps inner points of the unit circle (the limiting case of inner points G_n) onto inner points of G. Call these inner points the region G'. Therefore $\phi(w)$ maps the unit circle onto a region G' coincident with G, wholly or in part.

3. G' is a flush region.

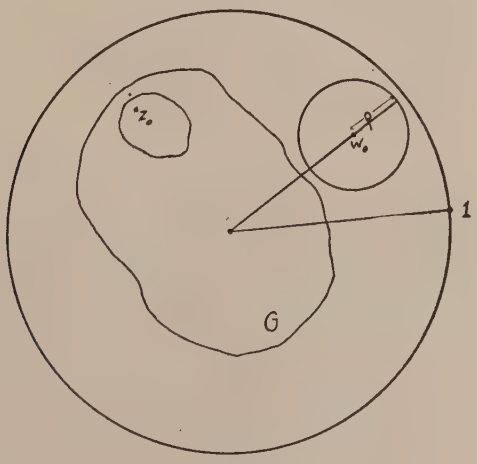


Fig. 8

Proof: To prove this it is necessary to establish the compactness of the suite of functions $f_n(z)$ (inverse to $\phi_n(w)$). They are defined for all values of z in G and map G onto G_n. Let w_o be any point within the unit circle. Then by 2^o, $z_o = \phi(w_o)$ is an inner point of G. With center w_o and radius ρ, $(\rho < 1 - |w_o|)$ draw a circle C (lying wholly in unit circle) and let w_i be any other point of C. $z = \phi(w)$ maps C onto a region which lies wholly in G and covers a part of G, and contains z_o. But if instead of using the limiting function $\phi(w)$, we map G_k onto G by $\phi_k(w)$, only that part of the circle C which lies in G_k will be mapped onto G.

However, from a definite $n > k$ onward, all regions G_n contain w_o (and the whole circle C for n sufficiently large).

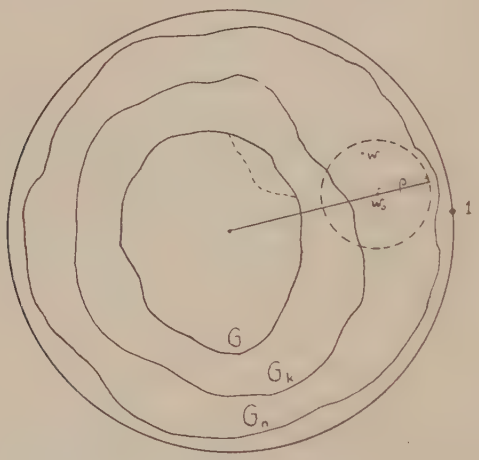


Fig. 9

For these "larger n" values, G contains z_o and points $z_i = \phi_n(w_i)$; all z_i for "n sufficiently large." The functions $f_n(z)$ of the inverse sequence are for all these sufficiently-large-n values defined at z_o and all points z_i. Now $\rho \geq |w_i - w_o|$; therefore $\rho \geq |f_n(z_i) - w_o|$. Since ρ can be taken arbitrarily small, from the first inequality, w_i differs from w_o by an arbitrarily small amount. Hence also z_i from z_o; hence, too, $f_n(z_i)$ from $f_n(z_o)$ (since all these functions are regular in the unit circle). Therefore, by the second inequality, $f_n(z_o)$ differs

from w_o by an arbitrarily small quantity (or $f_n(z)$ from w, since w_o is any point w within the unit circle). But since w may be arbitrarily close to the boundary of the unit circle, it is readily seen that the whole discussion "for n sufficiently large," and $|f_n(z)-w|\leq\rho$ implies: $\underset{n\to\infty}{\text{Lim}} f_n(z)$ exists and equals w, where w is the inverse of the limiting function $\phi(w)$. Hence the suite $f_n(z)$ is compact and its limit is $w=f(z)$.

Now it is evident that if of two functions (inverse one to the other) both are single valued for their respective bounded regions (maps one of the other), then both bounded regions are flush. The functions $\phi_n(w)$ (constructed after the pattern

$$\phi(w)=w\frac{A-w}{1-\bar{A}w}$$

given above) are single valued; the functions $f_n(z)$, inverses of $\phi_n(w)$ (of the form

$$f(z)=\tfrac{1}{2}\left[A+|A|z-\sqrt{(A+|A|z)^2-\frac{4|A|z}{A}}\right],$$

the negative sign before the radical determined by the condition $f(o)=o$), are also single valued. Since the functions of both suites are single valued independently of the value of n, the limiting functions $f(z)$ and $\phi(w)$ are single valued. Hence G' is flush.

4. G' is wholly coincident with G.

Proof: To prove this we use Koebe's theorem on distortion: "If $\psi(u)$ is a function finite and regular in the circle $|u|<\rho$, which maps this circle onto a flush region, and $\psi(o)=o$; $\psi'(o)=1$, then for any two points u_1 and u_2 of the circle $|u|\leq r<\rho$ there exists the inequality

$$Q(r)\leq\left|\frac{\psi'(u_1)}{\psi'(u_2)}\right|\leq\frac{1}{Q(r)}$$

where $Q(r)$ is some essentially positive real quantity $o<Q(r)<\infty$,

depending only on r and not at all on the particular form of the function ψ."

Apply this theorem to the functions $f_n(z)$. Imagine a circle entirely surrounding G. Since all $f_n(z)$ are defined for every z in G and map G onto a flush region G_n, we may, if we restrict $u_1 = z_1$ and $u_2 = z_2$ to the region G, apply Koebe's theorem to this region G. The origin is a point of G. Let $z_2 = o$ and $z_1 = z$ (any point in G). Since

$$f_n'(o) = 1/\phi_n'(o) \text{ and } \operatorname*{Lim}_{n\to\infty} \phi_n'(o) = 1 ,$$

then

$$\operatorname{Lim} f_n'(o) \equiv f'(o) = 1 .$$

Also

$$\operatorname*{Lim}_{n\to\infty} f_n'(z) = f'(z) .$$

Hence $m \leq |f'(z)| \leq M$, where m and M are two essentially real positive quantities differing from zero and infinity. From this inequality we get an upper and lower bound for the ratio of the lengths of a curve and its map. Let Z be a curve $z = z(t)$ having a derivative, and W a curve $w = f[z(t)]$, the map of Z. Their lengths call L_z' and L_w'; then

$$L_z' = \int_{z_0}^{z_1} |dz| = \int_{t_0}^{t_1} |z'(t)| dt \qquad (1)$$

$$L_w' = \int_{w_0}^{w_1} |dw| = \int_{t_0}^{t_1} |f'(z)|\ |z'(t)| dt . \qquad (2)$$

Since $m \leq |f'(z)| \leq M$, from (2) we get

$$m \int_{t_0}^{t_1} |z'(t)| dt < L_w' < M \int_{t_0}^{t_1} |z'(t)| dt .$$

Combining with (1) and dividing by L_z',

$$m < \frac{L_w'}{L_z'} < M .$$

These bounds having been established, the proof proceeds *per reductionem ad absurdum*. To prove G' wholly coincident with G, we suppose it is not. Construction: Place in G a region B wholly in G. Place in B a region B_1 wholly in B in such a way that B_1 contains in its interior inner points of G'. Assume a point a lying in both B_1 and G' arbitrarily close to boundary of G'. Draw a circle K, center a,

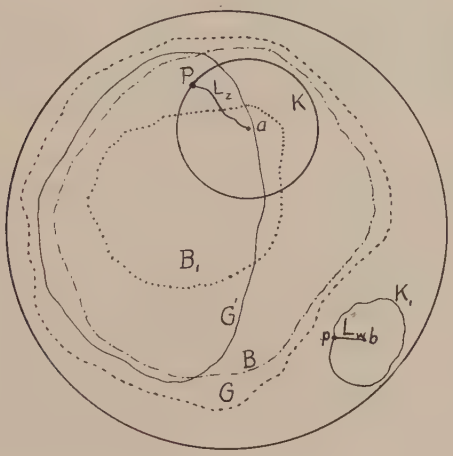


Fig. 10

lying wholly in B and containing points *outside* B_1 and G'. $f_n(z)$ maps a into some point b; circle K into some curve K_1. Pick out p on K closest to b and call the right line $|b-p|=L_w$. The point p maps by the inverse $\phi_n(w)$ into P on periphery of K and right line L_w into some curve $L_z \geq |a-P|$. The purpose of this construction is to insure a lower limit to $L_z \geq |a-P| > o$, since a is an inner point, and K contains outer points of B_1 and G'. Discussion: $f_n(z)$ for n sufficiently large will map some inner points of G' arbitrarily close to the circumference of the unit circle: ($\phi(w)$ maps unit circle onto $G' \leq G$). Let b, the correspondent of a, be one of these. Since K_1 lies wholly in the unit circle and p is by construction the nearest point on K_1 to b; and $|b-p| < L_w$ as n increases beyond all bounds,

$$f_n(z) \to f(z) \qquad L_w \to L'_w \qquad L_z \to L'_z$$

$$|b-p| \to o \qquad \text{while } |a-P| > o \text{ constantly .}$$

Therefore $L'_w \to 0$ and $L'_z > 0$. Hence the ratio $|L'_w/L'_z|$ does not lie between two essentially positive real numbers m and M differing from zero and infinity, contrary to Koebe's theorem. Therefore the supposition that G' is not wholly coincident with G is false. The function $\phi(w)$ consequently maps the unit circle on the whole flush surface G, and being holomorphic at all inner points of the unit circle, the mapping is everywhere conformal. Likewise $f(z)$ is regular at all inner points of G. Hence the mapping of G onto the unit circle is conformal at all points. Hence the theorem:

"There exists a function which maps any flush (or equivalently flush) region with at least two boundary points, one-to-one, conformally, and reversibly onto the flush interior of the unit circle."

It is scarcely to be expected that a non-flush surface will map onto the flush surface of a circle, one-to-one, conformally, reversibly. If the mapping is conformal at all inner points, the polygon (in Figure 6, page 29) and any other closed curve (crossing its boundary once and only once) will map into closed curves. On the flush surface of a circle, no closed curve can be drawn that crosses another closed curve once and only once. Hence the theorem cannot be extended, apparently, to non-flush surfaces.

VITA

Francis Joseph Gerst was born in Cincinnati, Ohio, June 29, 1882. His scholastic career is as follows: June, 1902, A.B., St. Xavier College, Cincinnati. (He entered the Society of Jesus September 2, 1902.) 1904–6, classics, education, history. 1906–9, philosophy, sciences, mathematics: A.M. (1908). 1910–11, physics: M.S. (1911). 1914–18, theology, canon law, Scripture—all at St. Louis University. (He was ordained 1917.) He spent 1920–21 at the University of Michigan, and 1921–23 at the Johns Hopkins University as a graduate student in mathematics, pure and applied, and in geophysics. Ill health preventing the completion of his work, the Board of University Studies kindly granted him leave to spend the last year *in absentia*. He wishes to express his sincere gratitude to the members of the Board and to all who have assisted him in his career: to his professors of mathematics: Father De Laak (St. Louis University); Professors Markley, Ziwet, Ford (University of Michigan); Dr. Musselman, Professors Murnaghan and Morley. He wishes especially to thank Professor Morley for his great kindness in inspiring and directing the work presented in these pages.